Statistics and Its Interface Volume 7 (2014) 3–7

In memoriam: Gang Zheng (May 6, 1965–January 9, 2014)

Nancy L. Geller* and Colin O. Wu

Figure 1. Dr. Gang Zheng.

The statistical community was deeply saddened by the death of our colleague, Gang Zheng, who lost his battle with head and neck cancer on Thursday January 9th. Gang received his BS in Applied Mathematics in 1987 from Fudan University in Shanghai. After serving as a teaching assistant at the Shanghai 2nd Polytechnic University, he emigrated to the U.S. in 1994 and received a master's degree in mathematics at Michigan Technological University in 1996. He then gained admission to the Ph.D. program in Statistics at The George Washington University and received his Ph.D. in 2000.

Immediately he joined the Office of Biostatistics Research at the National Heart, Lung and Blood Institute (NHLBI) of the National Institutes of Health (NIH), where he remained up to his death. From his interview seminar in early 2000, it was clear that the topic of his thesis, Fisher's Information and applications, was an area in which he could pursue research for many years. What was not obvious then was how prolific his research would become.

Over the past 13 years since he got his Ph.D., Gang collaborated with many researchers in developing statistical methods, including his colleagues at NHLBI, statisticians from other NIH institutes, and statistical faculty from universities in the US and other counties. He was one of the most productive researchers in biostatistics and statistics at NIH.

Gang developed new statistical procedures, which were motivated from his consultations at NHLBI, and published methodology papers, in which principal investigators (PIs) of NHLBI or NHLBI funded studies became his co-authors. One example is Zheng et al. (2005) [37], in which he developed new methods for sample size and power calculations for genetic studies taking into account the randomness of genotype counts given the allele frequency (the sample size and power are functions of the genotype counts). Dr. Elizabeth Nabel, the former Director of NHLBI, and her research fellow were co-authors on that paper. Another example is his consultation with Multi-Ethnic Study of Atherosclerosis (MESA) and Genetic Analysis Workshop (GAW16) with his colleagues Drs. Colin Wu, Minjung Kwak and Neal Jeffries. The studies contain data with outcome-dependent sampling and a mixture of binary and quantitative traits, for example, the measurements of a quantitative trait of all controls were not available. He developed a simple and practical procedure to analyze pleiotropic genetic association with joint binary (case-control) and continuous traits ([11], [44], [43]).

Most of Gang's research focused on three subject areas: (1) robust procedures and inference with nuisance parameters with applications to genetic epidemiology; (2) inference based on order statistics and ranked set sampling; and (3) pleiotropic genetic analysis with mixed trait data. Although he only started working on the last subject area in late 2012, he had already jointly published four papers in genetic and statistical journals ([18], [30], [28], [29]), and these results build a foundation for evaluating genetic data from combined big and complex studies.

His first paper in genetics dealt with applying robust procedures to case-control association studies ([6]). This paper has been cited over 160 times, according to the ISI Web of Science (Jan, 2014). It has become the standard robust test for the analysis of genetic association studies using a frequentist approach. The SAS JMP genomics procedure outputs the p-value of a robust test of Freidlin et al. (2002) [6] ([12]). Stephens and Balding (2009) [24] mentioned the lack of an analogous robust test of Freidlin et al. (2002) [6] for a Bayesian analysis. In 2010, an R package, RASSOC, for applying robust and usual association tests for genetic studies was developed by him and his coauthors ([31]).

In addition to novel applications of existing robust procedures to case-control genetic association studies, he developed several new robust procedures for genetic association studies. In Zheng and Ng (2008) [40], he and his coauthor

*Corresponding author.

used the information of departure from Hardy-Weinberg proportions to determine the underlying genetic model and incorporated genetic model selection into a test of association. Other robust procedures that he developed include Zheng et al. (2007) [41] on an adaptive procedure, Joo et al. (2009) [13] on deriving an asymptotic distribution for the robust test used by the Wellcome Trust Case-Control Consortium ([25]), and Kwak et al. (2009) [16] on robust methods in a two-stage procedure, so that the burden of genotyping can be reduced. Gang and his collaborators wrote an excellent tutorial on robust methods for linkage and association studies with the three most common genetic study designs ([14]). Kuo and Feingold (2010) [15] discussed several robust procedures developed by Gang and his collaborators, including Freidlin et al. (2002) [6] and Zheng and Ng (2008) [40], and compared the power of robust tests with other tests under various situations. So and Sham (2011) [23] reviewed and discussed many robust procedures developed by Gang, and also extended some of his procedures by allowing adjustment for covariates.

Gang developed an adaptive two-stage procedure for testing association using two correlated or independent test statistics with K. Song and R.C. Elston ([41]). His adaptive procedure was used by other researchers to design optimum multistage procedures for genome-wide association studies (e.g., [20], [27]). His use of two independent test statistics sequentially in Zheng et al. (2007) [41] was also used by others as one of the methods to replicate genetic studies ([19], [17]). Gang also wrote an important review article with R.C. Elston and D.Y. Lin on multistage sampling in human genetics studies ([5]).

In 2012, Dr. Zheng and his collaborators published a book entitled "Analysis of Genetic Association Studies" with Springer ([45]). It has over 436 pages with 40 illustrations. In the preface it states that "... both a graduate level textbook in statistical genetics and genetic epidemiology, and a reference book for the analysis of genetic association studies. Students, researchers, and professionals will find the topics introduced in Analysis of Genetic Association Studies particularly relevant. The book is applicable to the study of statistics, biostatistics, genetics and genetic epidemiology." Unlike other books in statistical genetics, Zheng et al. (2012) [45] also covers technical details and derivations that most other books omitted. In 13 years, Gang made a vast number of important contributions to statistical genetics.

In his early research (originating from on his Ph.D. thesis but extended considerably), Gang made important and extensive contributions to the computation and applications of Fisher information in order statistics and ordered data. In Zheng (2001) [32], he characterized the Weibull distribution in the scale-family of all life time distributions in terms of Fisher information contained Type II censored data and a factorization of the hazard function, which motivated further investigations by other researchers. For example, Hofmann, Balakrishnan and Ahmadi (2005) [8] extended his results using the Fisher information contained in the smallest order statistic. In a discussion paper by N. Balakrishnan ([1]), these results were also reviewed. Some of his work on Fisher information in order statistics has been extended to Fisher information in record values (e.g., [9]) and progressive censoring (e.g., [2]).

Gang studied where most Fisher information is located in samples from a location-scale family of distributions and provided theory and insight which explain why the tail and middle portions of the ordered data are most informative for the scale and location parameters, respectively. This added insight into an area initiated by the late John Tukey in the later part of the 1960's. Of particular interest is the fact that this is not true for the Cauchy distribution ([33], [35]). The latest version of the classical book "Order Statistics" 3rd ed. by H.A. David and H.N. Nagaraja (2003) [4] added a new section on Fisher information in order statistics (Sec. 8.2), which cites six papers Gang wrote on Fisher information in order statistics.

Applying his results, Sen et al. (2009) [22] proposed a novel study design for quantitative trait locus by oversampling the informative tails of the distribution identified in Zheng's papers. Ranked set sampling is a very useful alternative to random sampling, and still an active research area, but lacked applications beyond field studies or agriculture. Gang and his collaborators applied ranked set sampling to genetics association and linkage studies, which led to two important papers ([3], [36]). Their work motivated many further contributions from others, including David Clayton ([26]) and Danyu Lin ([10]).

A very important editorial contribution by Gang is his guest editorship for a special issue on statistical methods of genome-wide association studies for Statistical Science, co-edited with Prof. Jonathan Marchini and Dr. Nancy Geller ([38]). The special issue, which was published in November 2009, consists of 12 contributions from leading statisticians in the area. An introduction of this special issue appeared in the March 2010 IMS Bulletin ([39]). The three Editors were responsible for writing the proposal to the Editors of Statistical Science, identifying suitable contributors and getting their agreement to participate. The Executive Editor, David Madigan, of Statistical Science assigned Dr. Zheng to be the editor to handle the review process for all the submissions, except his own.

From his arrival, Dr. Zheng was a statistical consultant on the design and analysis of many NHLBI sponsored studies of cardiovascular diseases and asthma. One important project was the genetic study of in-stent restenosis, which started in 2004. With his colleagues Drs. Jungnam Joo (now at Korean National Cancer Center) and Nancy Geller, he designed this study, which was later expanded to the first genome-wide association study (GWAS) carried out by NHLBI in 2005, before NHLBI started funding GWAS. The original paper was published in Pharmacogenomics ([7]). In this study, he determined statistical procedures for quality control and developed methods for the analysis of the data. His early research

in GWAS earned him invitations to present his work at the 2007 JSM, at a seminar series of the Washington Statistical Society (2007), and at a seminar series at the Department of Biostatistics at the University of Pennsylvania (2008).

In 2004, Dr. Zheng became a statistical consultant for an NHLBI study: "A Case-Control Etiologic Study of Sarcoidosis" (ACCESS). A paper of ACCESS Research Group claimed that there was no association between immunoglobulin gene polymorphisms and sarcoidosis among African-Americans ([21]). A routine two-degree-of-freedom test built in SAS was applied to analyze the data by ACCESS investigators. He and his colleague developed a new efficiency robust procedure with constrained genetic models for the ACCESS data and re-analyzed the genetic association. They found it was statistically significant with the new procedure. The improvement came after incorporating the constraints on the genetic models but the routine chi-squared test ignores the restriction of the genetic model space. This research brought attention not only from the original PIs but also from the Steering Committee and the Data Safety and Monitoring Board of ACCESS. After more than six month discussions in several Steering Committee meetings and the consultation with a medical researcher outside of ACCESS, also under the pressure and objection from the original authors, the Steering Committee members finally voted to clear submission of Dr. Zheng's research for publication, which appeared in Statistics in Medicine ([42]). The ACCESS Research Group also decided to include this paper as an ACCESS publication. Dr. Lee Newman (Ex Officio of ACCESS and Professor of Medicine at Colorado School of Public Health) later invited Dr. Zheng to give a presentation based on his research findings.

When analyzing the data from his consultation for medical publications at NHLBI, Dr. Zheng not only developed more powerful statistical methods for the unique data, but also applied more appropriate tests to the data analysis. In one ongoing NHLBI intramural research to analyze association of candidate markers in osteoprotegerin with clinical phenotypes and its effects on cell biology in lymphangioleiomyomatosis, the original analyses were done by a staff scientist using some statistical tools built in Excel. Associations were tested using an allele-based test by comparing allele frequencies, and a genotype-based test by comparing genotype frequencies. Both results are reported. Although this is fine after correcting for multiple testing for two tests, Gang employed a newly developed method by him and his colleagues ([13]) to this dataset with the same allele-based and genotype-based tests but, instead of applying the Bonferroni correction for the two tests, he applied a more powerful approach to find p-values using the joint distribution of the two tests.

In addition to research contributions, Gang served as an Associate Editor of Statistics and Its Interface and co-edited several issues of the journal, the current one and an earlier one in honor of his thesis adviser Joe Gastwirth. He served as a referee for 43 journals and volumes, including JASA, Biometrics, Biometrika, Annals of Human Genetics, American Journal of Human Genetics and Statistics in Medicine.

Gang's degree of productivity is extremely rare and unusually versatile. He was honored for his work by election in 2005 as Fellow of the International Statistical Institute. He also gave a large number of invited talks, demonstrating the appreciation of his work by others.

One might think that such a productive researcher would be highly competitive. In fact, the opposite was true for Gang. He was an intellectually generous and nurturing colleague. He has mentored new members of the Office of Biostatistics Research at NHLBI both in research and collaboration. He has also mentored predoctoral fellows and served as a Ph.D. advisor to six students (two in China and four at George Washington University). In each case he published joint papers with these students. There was an old e-mail about one of them in which he said, "This is one of the things that makes me happy. This was a fine Ph.D. student. I gave him three topics for his Ph.D. thesis and he worked out five papers. I actually turned down authorship on the last two papers because I wanted him to come into my world and come out of it independently."

He has been equally generous to his other colleagues. We learned very quickly that if Gang asked you to collaborate with him on a research paper, to just say yes and be prepared to rearrange your own priorities so that you had time to work on it immediately, for the paper he was proposing would get written quickly, with or without your input. Indeed, Gang collaborated with almost all of his colleagues in the Office of Biostatistics Research. It was our pleasure to collaborate with him on nearly 20 papers between us. His efficiency and creativity were marvelous and inspiring. He was truly an intellectual leader in the Office of Biostatistics Research.

Gang also contributed admirably to the statistical profession by undertaking significant editorial responsibilities, serving on organizing and program committees of many meetings as well as organizing many sessions at various statistical meetings. He was also a member of the ASA Noether Award Committee. These many activities illustrate Gang's generosity as a colleague and his dedication to the profession. Despite the setback of his illness, he continued to be highly productive and published seven new papers in 2013.

Gang's efficiency, creativity and generosity were truly inspiring. Those of us who have been his colleagues and collaborators will always remember the experience. He will be sorely missed.

Received 29 January 2014

REFERENCES

[1] Balakrishnan, N. (2007). Progressive censoring methodology: an appraisal. *Test* **16**, 211–259. MR2393645

[2] Balakrishnan, N., Burkschat, M., Gramer, E. and Hoffman, G. (2008). Fisher information based progressive censoring plans. *Computational Statistics and Data Analysis* **53**, 366–380. MR2649092

[3] Chen, Z., Zheng, G., Ghosh, K. and Li, Z. (2005). Linkage disequilibrium mapping for quantitative trait loci by selective genotyping. *American Journal of Human Genetics* **77**, 661–669.

[4] David, H. A. and Nagaraja, H. N. (2003). *Order Statistics, Third Edition*. Wiley, Hoboken, New Jersey. MR1994955

[5] Elston, R. C., Lin, D. Y. and Zheng, G. (2007). Multi-stage sampling for genetic studies. *Annual Reviews of Genomics and Human Genetics* **8**, 327–342.

[6] Freidlin, B., Zheng, G., Li, Z. and Gastsirth, J. L. (2002). Trend tests for case-control studies of genetic markers: Power, sample size and robustness. *Human Heredity* **53**, 146–152.

[7] Ganesh, S. K., Skelding, K. A., Mehta, L., O'Neill, K., Joo, J., Zheng, G., Goldstein, J., Simari, R., Billings, E., Geller, N. L., Holmes, D., O'Neill, W. W. and Nabel, E. G. (2004). Rationale and study design of the CardioGene study: genomics of in-stent restenosis. *Pharmacogenomics* **5**, 949–1004.

[8] Hofmann, G., Balakrishnan, N. and Ahmadi, J. (2005). Characterization of hazard function factorization by Fisher information in minima and upper record values. *Statistics and Probability Letters* **72**, 51–57. MR2126293

[9] Hofmann, G. and Nagaraja, H. N. (2003). Fisher information in record data. *Metrika* **57**, 177–193. MR1969251

[10] Huang, G. B. E. and Lin, D. Y. (2007). Efficient association mapping of quantitative trait loci with selective genotyping. *American Journal of Human Genetics* **80**(3), 567–576.

[11] Jeffries, N. and Zheng, G. (2009). Evaluation of an optimal receiver operating characteristics procedure. *BMC Proceedings* **15**;3 Suppl 7:S56.

[12] *JMP Life Science User Manual (Genomics and Clinical)* (2014). SAS Institute Inc., Cary, North Carolina.

[13] Joo, J., Kwak, M., Ahn, K. and Zheng, G. (2009). A robust genome-wide scan statistic of the Wellcome Trust Case-Control Consortium. *Biometrics* **65**, 1115–1122. MR2756499

[14] Joo, J., Kwak, M., Chen, Z. and Zheng, G. (2010). Efficiency robust statistics for genetic linkage and association studies under genetic model uncertainty. *Statistics in Medicine* **29**, 158–180. MR2751387

[15] Kuo, C. L. and Feinberg, E. (2010). Letter to the Editor. *Genetic Epidemiology* **34**, 772.

[16] Kwak, M., Joo, J. and Zheng, G. (2009). A robust test for two-stage design in genome-wide association studies. *Biometrics* **65**, 1288–1295. MR2756517

[17] Laird, N. M. and Lange, C. (2009). The role of family-based designs in genome-wide association studies. *Statistical Science* **24**, 388–397. MR2779333

[18] Li, Q., Hu, J., Ding, J. and Zheng G. (2014). Fisher's method combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations. *Biostatistics* Oct 29, 2013 [Epub ahead of print].

[19] Murphy, A., Weiss, S. T. and Lange, C. (2008). Screening and replication using the same data set: testing strategies for family-based studies in which all probands are affected. *PLOS Genetics* DOI: 10.1371/journal.pgen.1000197.

[20] Pahl, R., Schafer, H. and Muller, H. H. (2009). Optimal multistage designs – a general framework for efficient genome-wide association studies. *Biostatistics* **10**, 297–309.

[21] Pandey, J. P., Frederick, M. and Access Research Group. (2002). TNF-α, IL1-β, and immunoglobulin (GM and KM) gene polymorphisms in sarcoidosis. *Human Immunology* **63**, 485–491.

[22] Sen, S., Johannes, F. and Broman, K. W. (2009). Selective genotyping and phenotyping strategies in a complex trait context. *Genetics* **181**, 1613–1626.

[23] So, H. and Shan P. C. (2011). Robust association tests under different genetic models, allowing for binary or quantitative traits and covariates. *Behavior Genetics* **41**, 768–775.

[24] Stephens, M. and Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics* **10**, 681–690.

[25] The Welcome Trust Case Control Consortium (WTCCC). (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 1753–1773.

[26] Wallace, C., Chapman, J. M. and Clayton, D. G. (2006). Improved power offered by a score test for linkage disequilibrium mapping of quantitative-trait loci by selective genotyping. *American Journal of Human Genetics* **78**(3), 498–504.

[27] Won, S. and Elston, R. C. (2008). The power of independent types of genetic information to detect association in a case-control study design. *Genetic Epidemiology* **32**, 731–756.

[28] Wu, C. O., Zheng, G. and Kwak, M. (2013). A joint regression analysis for genetic association studies with outcome stratified samples. *Biometrics* **69**, 417–426. MR3071060

[29] Xu, J., Zheng, G. and Yuan A. (2013). Case-control genome-wide joint association study using semiparametric empirical model and approximate Bayes factor. *Journal of Statistical Computing and Simulation* **83**, 1191–1209.

[30] Yan, T., Li, Q., Li, Y., Li, Z. and Zheng, G. (2013). Genetic association with multiple traits in the presence of population stratification. *Genetic Epidemiology* **37**, 571–580.

[31] Zang, Y., Fung, W. K. and Zheng, G. (2010). Simple algorithms to calculate asymptotic null distributions for robust tests in case-control genetic association studies in R. *Journal of Statistical Software* **33**(8).

[32] Zheng, G. (2001). A characterization of the factorization of hazard function by the Fisher information under Type II censoring with application to Weibull distribution. *Statistics and Probability Letters* **52**, 249–253. MR1838212

[33] Zheng, G. and Gastwirth, J. L. (2000). Where is the Fisher information in an ordered sample? *Statistica Sinica* **10**, 1267–1280. MR1804545

[34] Zheng, G. and Gastwirth, J. L. (2001). On the Fisher information in randomly censored data. *Statistics and Probability Letters* **52**, 421–426. MR1841610

[35] Zheng, G. and Gastwirth, J. L. (2002). Do tails of symmetric distributions contain more Fisher information about the scale parameter? *Sankhya Series B* **64**, 289–300. MR1993915

[36] Zheng, G., Ghosh, K., Chen, Z. and Li, Z. (2006). Extreme rank selection for linkage analysis of quantitative trait loci using selected sib-pairs. *Annals of Human Genetics* **70**, 857–866.

[37] Zheng, G., Joo, J., Ganesh, S., Nabel, E. and Geller, N. L. (2005). On averaging power for genetic association and linkage studies. *Human Heredity* **59**, 14–20.

[38] Zheng, G., Marchini, J. and Geller, N. L. (2009). Introduction to the Special Issue: Genome-wide association studies. *Statistical Science* **24**, 387. MR2779332

[39] Zheng, G., Marchini, J. and Geller, N. L. (2010). Genome-wide association studies. *IMS Bulletin* **39**(2), 10.

[40] Zheng, G. and Ng, H. K. T. (2008). Genetic model selection in two-phase analysis for case-control association studies. *Biostatistics* **9**, 391–399.

[41] Zheng, G., Song, K. and Elston, R. C. (2007). Adaptive two-stage analysis of genetic association for case-control designs. *Human Heredity* **63**, 175–186.

[42] Zheng, G. and Tian, X. for ACCESS Research Group (2006). Robust trend tests for genetic association using matched case-control design. *Statistics in Medicine* **25**, 3160–3173. MR2252289

[43] Zheng, G., Xu, J., Yuan, A. and Wu, C. O. (2013). Impact on modes of inheritance and relative risks using extreme sampling when designing genetic association studies. *Annals of Human Genetics* **77**, 80–84.

[44] Zheng, G., Wu, C. O., Kwak, M., Jiang, W., Joo, J. and Lima, J. A. C. (2012). Joint analysis of binary and quantitative trait with data sharing and outcome-dependent sampling. *Genetic Epidemiology* **36**, 263–273.

[45] Zheng, G., Yang, Y., Zhu, X. and Elston R. C. (2012). *Analysis of Genetic Association Studies*. Springer, New York. MR2895171

Nancy L. Geller
Office of Biostatistics Research
National Heart, Lung and Blood Institute
6701 Rockledge Drive
Bethesda, MD 20892-7913
USA
E-mail address: gellern@nhlbi.nih.gov

Colin O. Wu
Office of Biostatistics Research
National Heart, Lung and Blood Institute
6701 Rockledge Drive
Bethesda, MD 20892-7913
USA
E-mail address: wuc@nhlbi.nih.gov